56

E49

(19) **JAPANESE PATENT OFFICE**

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **11327995 A**

(43) Date of publication of application: **30.11.99**

(51) Int. Cl	G06F 12/00
	G06F 13/00
	G06F 17/40
	H04L 12/54
	H04L 12/58

(21) Application number: 10135195

(22) Date of filing: 18.05.98

(71) Applicant: NIPPON TELEGR & TELEPH CORP <NTT>

(72) Inventor: YOKOMICHI SEIJI
MIURA NOBUYUKI
TAKAHASHI KATSUMI
SHIMA KENICHI
WASHISAKA KOICHI

(54) **METHOD AND DEVICE FOR GATHERING INFORMATION AND RECORDING MEDIUM HAVING RECORDED INFORMATION GATHERING PROGRAM**

(57) Abstract:

PROBLEM TO BE SOLVED: To make predictable resources which need to be gathered without actually gathering them by classifying resources decentralized over a network for every content and selectively and automatically gathering only necessary resources.

SOLUTION: A worldwide web(WWW) document of data saved in a data saving part 4 is supplied to a content analysis part 11 through a hyperlink extraction part 5 and a hyperlink evaluation part 6. The content analysis part 11 takes a morpheme analysis of the WWW document to extract nouns, proper nouns, and undefined words, and generates word lists by documents and supplies them to a content prediction and learning part 12. Featured patterns of specific fields are extracted from WWW documents and their lists are generated by the documents and supplied to the content prediction and learning part 12. Then HTML tags are extracted from a WWW document and a list of tags for hyperlinks is generated and supplied to the content prediction and learning part 12. Thus, resources decentralized over the network are classified for every content and only necessary resources are gathered.

COPYRIGHT: (C)1999,JPO

-1-